Year: 2017

A crowdsourced model of landscape preference

Chesnokova, Olga ; Nowak, Mario ; Purves, Ross S

Abstract: The advent of new sources of spatial data and associated information (e.g. Volunteered Geographic Information (VGI)) allows us to explore non-expert conceptualisations of space, where the number of participants and spatial extent coverage encompassed can be much greater than is available through traditional empirical approaches. In this paper we explore such data through the prism of landscape preference or scenicness. VGI in the form of photographs is particularly suited to this task, and the volume of images has been suggested as a simple proxy for landscape preference. We propose another approach, which models landscape aesthetics based on the descriptions of some 220000 images collected in a large VGI project in the UK, and more than 1.5 million votes related to the perceived scenicness of these images collected in a crowdsourcing project. We use image descriptions to build features for a supervised machine learning algorithm. Features include the most frequent uni- and bigrams, adjectives, presence of verbs of perception and adjectives from the "Landscape Adjective Checklist". Our results include not only qualitative information relating terms to scenicness in the UK, but a model based on our features which can predict some 52% of the variation in scenicness, comparable to typical models using more traditional approaches. The most useful features are the 800 most frequent unigrams, presence of adjectives from the "Landscape Adjective Checklist" and a spatial weighting term.

DOI: https://doi.org/10.4230/LIPIcs.COSIT.2017.19

Originally published at:
Chesnokova, Olga; Nowak, Mario; Purves, Ross S (2017). A crowdsourced model of landscape preference. In: COSIT 2017: International Conference on Spatial Information Theory, L'Aquila (I), 4 September 2017 - 8 September 2017, Dagstuhl Publishing.
DOI: https://doi.org/10.4230/LIPIcs.COSIT.2017.19

A Crowdsourced Model of Landscape Preference

Olga Chesnokova[1], Mario Nowak[2], and Ross S. Purves[3]

1 **Department of Geography, University of Zürich, Zürich, Switzerland**
`olga.chesnokova@geo.uzh.ch`
2 **Department of Geography, University of Zürich, Zürich, Switzerland**
`mario.h.nowak@gmail.com`
3 **Department of Geography, University of Zürich, Zürich, Switzerland**
`ross.purves@geo.uzh.ch`

Abstract

The advent of new sources of spatial data and associated information (e.g. Volunteered Geographic Information (VGI)) allows us to explore non-expert conceptualisations of space, where the number of participants and spatial extent coverage encompassed can be much greater than is available through traditional empirical approaches. In this paper we explore such data through the prism of landscape preference or *scenicness*. VGI in the form of photographs is particularly suited to this task, and the volume of images has been suggested as a simple proxy for landscape preference. We propose another approach, which models landscape aesthetics based on the descriptions of some 220000 images collected in a large VGI project in the UK, and more than 1.5 million votes related to the perceived scenicness of these images collected in a crowdsourcing project. We use image descriptions to build features for a supervised machine learning algorithm. Features include the most frequent uni- and bigrams, adjectives, presence of verbs of perception and adjectives from the "Landscape Adjective Checklist". Our results include not only qualitative information relating terms to scenicness in the UK, but a model based on our features which can predict some 52% of the variation in scenicness, comparable to typical models using more traditional approaches. The most useful features are the 800 most frequent unigrams, presence of adjectives from the "Landscape Adjective Checklist" and a spatial weighting term.

1998 ACM Subject Classification I.7.0 Document and Text Processing

Keywords and phrases VGI, crowdsourcing, semantics, landscape preference

Digital Object Identifier 10.4230/LIPIcs.COSIT.2017.19

1 Introduction

The advent of new sources of spatial data, and in particular those which are generated not through a top-down, regulated process, but bottom-up, by individuals with varying backgrounds and motivations, has brought with it new opportunities for research. In particular, the advent of spatial data associated with natural language, typically in the form of tags or unstructured text provide a potential route to exploring ways in which space is described in language, albeit typically in corpora where we as researchers have very little control. The data studied in such research can be produced in a number of ways, and differing, but overlapping, definitions have been assigned to such data including those related to volunteered geographic information (VGI), crowdsourcing, user-generated content, social media, citizen science and so on [7]. These definitions are important since they have implications for the ways in which data are produced, and in turn the ways in which they can reasonably be interpreted.

13th International Conference on Spatial Information Theory (COSIT 2017).
Editors: Eliseo Clementini, Maureen Donnelly, May Yuan, Christian Kray, Paolo Fogliaroni, and Andrea Ballatore; Article No. 19; pp. 19:1–19:13
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

One obvious, and much studied, source of such data are the tags and descriptions associated with georeferenced images. Here, researchers typically assume that images and their descriptions often capture information about named locations, their properties and, occasionally, notions related to sense of place (e.g. [28, 16]). Indeed, Fisher and Unwin [8] presciently recognised this potential in 2005, stating that "GI theory articulates the idea of absolute Euclidean spaces quite well, but the socially-produced and continuously changing notion of place has to date proved elusive to digital description except, perhaps, through photography and film. (p. 6)." Nonetheless, in practice analysing text and extracting information related to place has proved challenging, and many studies have either focussed above all on exploring the properties of text related to location, with limited or no opportunities for validation, or on using counts of images as a proxy for some spatially varying phenomena and generating appropriate statistical models (e.g. [5, 36, 37]).

The act of georeferencing images typically implies that an individual wishes to relate a particular image to an event (not relevant in the context of this work) or a location. The act of producing an image however is not random, and neither is the act of choosing to share an image with others in an online source [11]. Images capturing locations presumably capture perceptually salient elements of a landscape, and thus, accompanied by their descriptions might provide us with clues as to how landscape is conceptualised and parcelled up into cognitive entities [22]. Understanding landscape, and the ways in which it is perceived is not merely an abstract research question, but one with considerable direct policy and societal relevance, since landscapes are the subject of national and international policies and regulation. Contemporaneously with the emergence of new data sources such as those described above, has been an increasing realisation in many areas of policy that there is a need to include not only top-down definitions of landscapes in policy work, but also to capture bottom-up ways in which landscapes are perceived and experienced. Even seemingly simple notions such as landscape aesthetics have proved remarkably challenging to generalise and model spatially, and although methods based in the social sciences can capture well the diversity of opinions about individual locations, they are ill-suited to characterising large regions [37].

In this paper we set out to demonstrate, through the use of two, related, datasets, how we can firstly, capture through textual descriptions, elements of a landscape which are perceived as more or less attractive across a large region. To do so, we combine descriptions of georeferenced images which are an excellent example of VGI *sensu* Goodchild [12] with a large crowdsourced data containing *scenicness* rating for more than 220000 images. We then develop and evaluate a predictive model of scenicness, which as its primary input uses text describing images, and thus aims to model scenicness as a function of language.

1.1 Related work

In the following we briefly set out related work from two key areas. Firstly, we summarise concepts related to landscape aesthetics and its assessment. Secondly, we explore examples of research which have used novel data sources to explore landscape properties in a range of ways.

Theories seeking to explain landscape perception and aesthetics typically focus on both evolutionary and cultural influences [19, 15]. Evolutionary approaches assume that preferences with respect to landscape relate to the ability of landscapes to meet human needs such as 'prospect' (i.e. the ability to command a landscape through sight) and 'refuge' (the potential to conceal oneself in a landscape) [1]. Other, related concepts include the ability to 'make sense' of the environment (coherence and legibility of landscapes), and 'involvement' or ability to function well in the environment (complexity and mystery of landscapes) [18]. Cultural influences on landscape preference are recognised in the emergence of work on landscape and language, for example, through the study of ethnophysiography [22] which notes the importance of cultural influences and the absence of universally shared landscape elements.

Irrespective of the theoretical perspective taken, typical approaches to capturing landscape perception have focussed on in-situ methods using, for example, interviews and participatory mapping [2, 27]. However, the need to be on site makes such approaches poorly suited to capturing dynamic landscape preferences over large areas, and also makes it difficult to control potential influences. Such limitations, and the simple need to generate more lab-based reproducible experiments, led to the development of approaches based around photographs of landscapes where participants can be presented with images controlling the visual field [31], seasonal changes, or introducing extra factors (e.g. presence of animals [17] or anthropogenic objects [20]).

The advent of VGI, and the realisation that such data might contain diverse, independent and decentralised information, provided opportunities to replicate previous work on geographic concepts [34], and to demonstrate that such data were a reliable source of information about landscape characteristics and the ways in which landscapes were categorised [6, 28]. In parallel, the need to generate landscape indicators related to cultural ecosystem services and landscape preferences over large areas has led some of researchers to use the position and number of images taken as a proxy indicator of landscape preference [36, 37], or to incorporate the number of individuals taking pictures [3, 11] and their origins [10]. Others have realised that the images themselves contain information central to understanding landscape preference, and have analysed image content to explore cultural ecosystem services [30]. The importance of scenicness in a policy context, and the possibilities offered by new data sources are recognised in recent work exploring the link between wellbeing and scenicness using crowdsourced data, and attempting to model scenicness using user generated content [33, 32].

In this paper we seek to build on previous work in two key ways. Firstly, in-situ and lab-based studies of landscape preference have typically worked, of necessity, with relatively small groups of participants in focussed, often coherent, landscapes. Our study, by using VGI at the scale of Great Britain, allows us to explore landscape preferences across a whole country, and to explore regional differences between such preferences. Secondly, attempts to model scenicness have typically focussed on using spatial data in some form as explanatory variables (for example number of images, elevation, number of visible pixels, landcover type, etc.). We take an approach which we argue is likely to be closer to the way in which a particular landscape is perceived, and build a model of scenicness which uses language (in the form of words and phrases extracted from written descriptions) as explanatory variables.

In the following, we first describe the datasets on which we carried out our experiment, and the steps we took in processing, analysing and modelling scenicness with these data. We then present our results, demonstrating that the words used to describe scenic areas make clear distinctions especially between scenes perceived to be more or less anthropogenically influenced. Our model of scenicness is capable of explaining about 52% of the variance in scenicness in space, which is comparable to typical state of the art approaches. We then discuss the implications of these results, before concluding with some suggestions for future research.

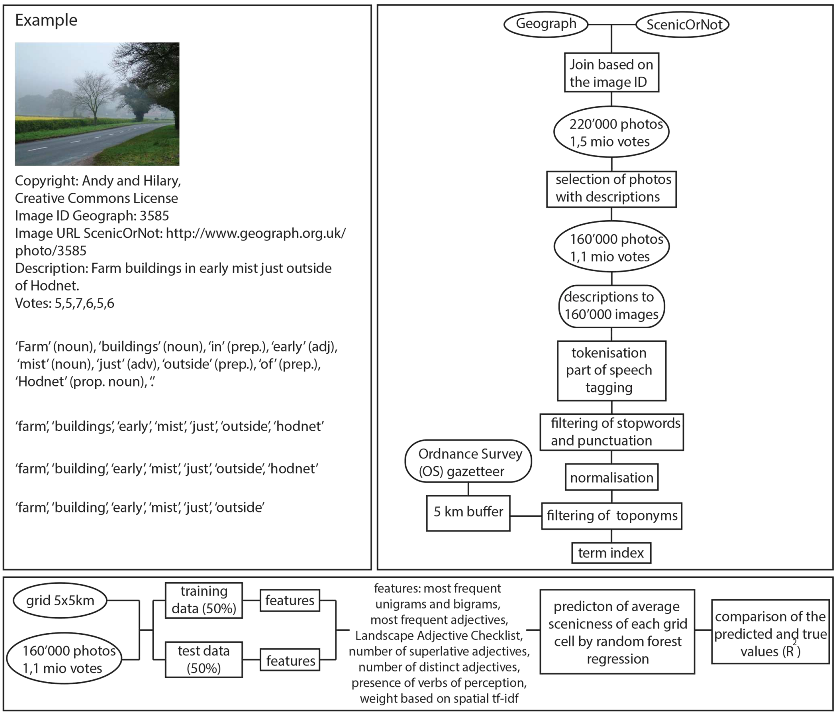


Figure 1 Steps of the data acquisition and preprocessing with an example.

2 Data and methods

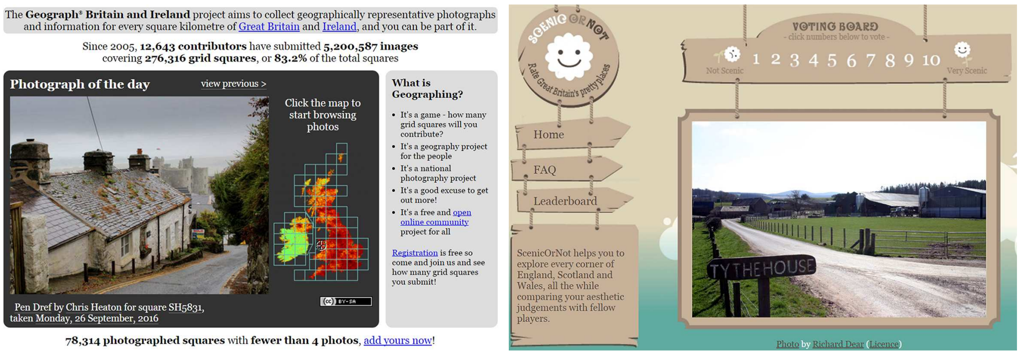
As set out above, our aims are twofold. Firstly, we wish to identify which terms are typically used with more or less scenic images, as described by votes in ScenicOrNot project and, secondly, based only on terms describing images to develop a spatially contiguous model of scenicness at the country level. In the following we describe the datasets used, and in particular aspects relevant to our work. We then set out our approach to processing the corpus, before describing the features used in producing our spatial model of scenicness. Fig. 1 gives a visual overview of the material which follows.

2.1 Data and study region

We use two unique, and related, datasets in this work. The Geograph[1] dataset (Fig. 2a) is a crowdsourced collection with more than 12000 contributors, launched in 2005, with the aim of collecting "geographically representative photographs and information for every square kilometre of Great Britain and Ireland." The project takes the form of a game, with users receiving points for uploading georeferenced images and associated descriptions, and content is moderated. The entire dataset is available under a Creative Commons Licence, and in this paper we used a version downloaded in June 2016 consisting of ca. five million images.

The ScenicOrNot[2] project (Fig. 2b) was initiated in 2009 by MySociety and is currently hosted by the Data Science Lab at Warwick Business School. The goal of the project is

[1] http://www.geograph.org.uk/
[2] http://scenicornot.datasciencelab.co.uk/



(a) Interface of the Geograph project

(b) Interface of the ScenicOrNot project

Figure 2 Interface of the Geograph project (Copyright Chris Heaton, Creative Commons Licence) and the ScenicOrNot project, where users can rate a Geograph photograph from 1 (not scenic) to 10 (very scenic).

to crowdsource scenicness ratings using Geograph images. In contrast to Geograph, where it is reasonable to assume that users uploading images typically also took the pictures in question (and thus visited the landscape), the ScenicOrNot project is purely internet based. Participants, about whom no demographic information is collected, are presented with a series of random images, with neither associated locations or descriptions, and asked to rate them on a scale of 1 (not scenic) to 10 (scenic) for scenicness. More than 220000 Geograph images had amassed some 1.5 million votes by June 2016 in the ScenicOrNot collection.

In the following our corpus consists of the 160000 Geograph images which both have a description, and are associated with three or more votes in ScenicOrNot.

2.2 Corpus processing

Our aim in corpus processing was to explore how terms used in describing Geograph images were associated with scenicness ratings. Since our starting point are natural language captions, standard corpus processing steps were applied. In the following, we briefly describe these steps, which were, in the main, carried out using the Python-based NLTK[3] library.

Each image description was in parallel tokenised, and part of speech tagged. The tokens were then filtered for stopwords and punctuation, before being normalised by changing all tokens to lower case and reducing tokens to their lemmas. Our aim was to build a term index, with associated features, for use in exploring the semantics of scenic locations.

Since we were explicitly not interested in the names of locations, we filtered toponyms from descriptions using gazetteer look-up in a 5km window around the coordinates associated with images. We used a freely available gazetteer, based on the 1:50000 maps from the Ordnance Survey for this process. This approach aims to strike a balance between removing local toponyms, which may be the subject of considerable semantic ambiguity (e.g. does bath refer to a place to bathe or the historic city) and retaining tokens which are being used in a non-toponymic sense.

Having performed these steps we are left with a term index, where unique entries are made up of tuples containing normalised tokens (unigrams and bigrams) present once or more in a description, part of speech tagging and the images IDs with which they are associated. Since

[3] http://www.nltk.org/

each term can be present in one or more images, and each image is ranked three or more times, we assign an average scenicness to every term in our index. Importantly, identical tokens having different parts of speech will have different values of average scenicness. Furthermore, since we store image IDs, we also have access to all locations associated with a term, the array of votes and an overall frequency of the term, based on the number of images described using a given term. Using our term index, it is possible to generate lists of terms, ranking or filtering by, for example, average scenicness, part of speech or frequency.

2.3 Feature choice and modelling scenicness

The final step in our approach was to create a spatially contiguous model of scenicness based on our term index. We predict scenicness for 5km grid cells, using Random Forests regression, which is a state of the art non-linear, non-parametric method in supervised machine learning, and which requires no assumptions with respect to the data distribution [4]. Our choice of 5km was motivated by the underlying 1km granularity of the Geograph data and its associated spatial distribution. We report briefly on sensitivity to resolution in the discussion.

A key task in creating such a model is the choice of appropriate features. Our basic approach was to use training data associated with 5km grid cells, where average scenicness was associated with features based on our terms. Only descriptions consisting of at least five tokens, after filtering as described above, were used in the model. The simplest possible feature set would be one based purely on unigrams, that is to say individual tokens from image descriptions found in grid cells (e.g. ‘hill’, ‘mountain’, ‘shop’, etc.).

However, in natural language processing [21] it is typical to also consider n-grams, and here we also experimented with bigrams (e.g. sequences of two tokens such as ‘steep hill’, ‘rugged mountain’, ‘closed shop’) as features. By reducing the feature space it is often possible to maintain model predictive capacity, while improving performance, and we also experimented by reducing the number of unigrams considered to the n-most frequent. Other features of our data, and previous work on landscape description, suggest additional potential model features which are listed below:

- adjectives alone: since adjectives are assumed to be strong indicators of subjectivity and sentiment; [14], we used unigrams consisting only of frequent adjectives;
- “Landscape Adjective Checklist”: presence of adjectives pertaining specifically to landscape in Craik’s list [24];
- the number of superlative adjectives as identified during part of speech tagging, with the assumption that superlatives are more likely to be used in more scenic areas;
- the number of distinct adjectives found in a description, with the assumption that more adjectives are used in more scenic areas;
- the presence of a verb of perception [39], where we assume that the presence of verbs of perception may indicate descriptions more relevant with respect to scenicness (e.g. by reducing the weight of descriptions focussing on historical events at a location);
- a weight based on spatial tf-idf [29]: here terms which are used frequently in an individual grid cell, but rarely in the collection as a whole are given a higher weight.

2.4 Training and test data

In any supervised model it is necessary to generate both training and test datasets. However, the way in which the data are split can have important implications for not only the quality of the model, but also for any implications which can be drawn from the results. Since an important property of crowdsourced data are user-generated biases in data production [13], we

(a) Scenicness between 1–3,
$n_{images} = 14072$, $n_{users} = 2137$.

(b) Scenicness between 3–5,
$n_{images} = 155822$, $n_{users} = 4170$.

(c) Scenicness between 5–7,
$n_{images} = 79752$, $n_{users} = 3340$.

(d) Scenicness between 7–10,
$n_{images} = 3134$, $n_{users} = 851$.

Figure 3 Average scenicness for 150 most frequent nouns extracted from image descriptions - font size indicates relative frequency within scenicness range.

considered these, as well as the desired spatial contiguity of our model in generating training and test data. Thus, our models were trained (and tested) on the following configurations, with 50% assigned to training and test data respectively in both cases:

- fully random: image descriptions are simply selected at random from the full corpus;
- user dependent random: since we expect individual users to write characteristic descriptions, and since Geograph is subject to participation inequality, meaning that a single user may contribute a large proportion of the descriptions in a single area, we select random images while allowing individual users only to appear in either training or test datasets.

3 Results and interpretation

3.1 Semantics of scenicness

The word clouds in Fig. 3 exemplify our results, illustrating the average scenicness of nouns after part of speech tagging of image descriptions. A number of features are worthy of observation here. At a first glance, the lowest rated scenicness values are related to nouns which are clearly in developed areas (e.g. ‘motorway’, ‘housing’, ‘shop’, ‘stadium’). The highest rated scenicness nouns include Gaelic words, terms related to natural processes, wildlife and some esoteric examples (e.g. ‘coire’, ‘avalanche’, ‘otter’, ‘backcloth’). However, these classes contain a small proportion of the total set (ca. 6%), with only some 1% of nouns being found in the most scenic class. Thus, many of these nouns belong to the long tail of our data, and although they reflect a clear split between developed areas and more natural landscapes (associated with Gaelic placenames in the Highlands of Scotland) we should be careful not to overinterpret these terms.

Unsurprisingly, since each image was rated at least three times, and many of the nouns are associated with multiple images, the vast majority (94%) of nouns have average scenicness ratings of between 3 and 7. Exploring these classes, it becomes apparent that the clear split so visible in the two extreme classes is much less prominent. Thus, we find that nouns such as 'village', 'lane' and 'wood' are all rated on average 3–5, even though these might be terms typically expected to be associated with more rural, and thus potentially more scenic images. However, exploring the nouns rated 5–7 it again becomes clear that differences exist. Here, many more nouns appear to relate to perceived natural (as opposed to rural) scenes (e.g. 'moorland', 'summit', 'ridge').

3.2 Predicting scenicness

We tested the goodness of fit of our Random Forest regression using the features as described above, and two different configurations of test and training data. Independent of the configuration chosen, we only predicted scenicness values for grid cells where at least two descriptions were present in both training and test data.

Goodness of fit improved as we increased the number of unigrams in the model until we reached the 800 most frequent unigrams. Including presence of adjectives from the "Landscape Adjective Checklist" by Craik and weighting according to spatial tf-idf further increased goodness of fit to a maximum value of around 52% (52.4% in the case of fully random and 52.0% in the case of user dependent random division on training and test data).

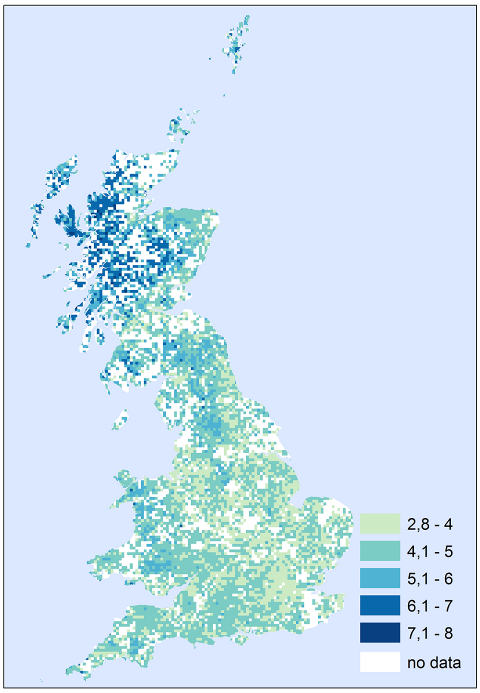
Fig. 4 shows the spatial pattern of predicted scenicness for both configurations. Particularly evident here are the larger number of grid cells for which no value could be predicted where training and test data were randomly selected according to users. Here, the effects of participation inequality result in many grid cells where the majority of images and associated descriptions were taken by a single user, and we thus cannot predict scenicness. However, given the limited variation in model goodness of fit, it appears that this restriction may be unnecessary.

A further important issue in our model is the existence of spatial autocorrelation in model residuals. Testing for Morans-I revealed values of around 0.12 according to model configuration, implying that the chances of random clustering in our model are less than 1%. A typical approach to assessing the influence of spatial autocorrelation in Random Forest regression is therefore to include grid centroids as features in the model [23]. Doing so increased goodness of fit to 56% and reduced spatial autocorrelation in the residuals to 0.05. An alternative model including spatial information by assigning county names (administrative units) to every image, resulted in a decrease of Morans's I to 0.10, with goodness of fit remaining at 52%. This approach includes local neighbourhood relationships and more natural divisions of landscape (since at least in the UK county boundaries typically are a mix of the *fiat* and *bona fide*). Since model results for a model based only on language and containing additional explicit spatial information are similar we thus conclude that our results are not biased by spatial autocorrelation [37].

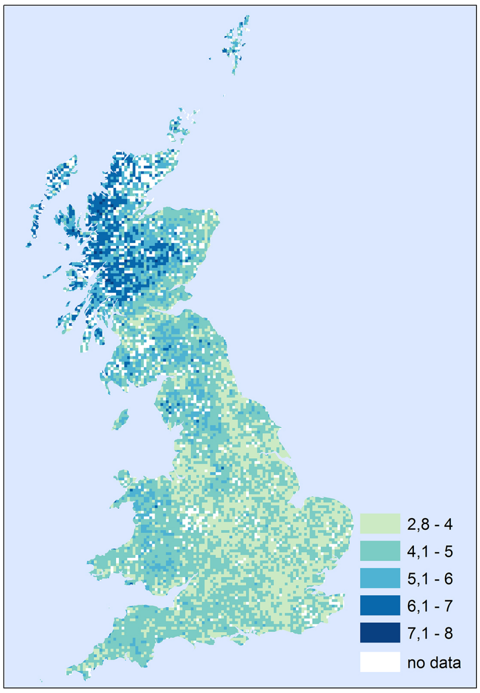
4 Discussion

In this paper we explored the use of two, related datasets which were both generated by the crowd, though in very different ways, to understand how landscape, and in particular scenicness is captured in language.

Our results were generated after a typical natural language pipeline to tokenise, classify and filter image descriptions. Importantly, we also included a step to remove toponyms from image descriptions, since we were not interested in the names of scenic places, but rather in their properties. Our results demonstrate a clear transition from nouns associated with



(a) User-based division.



(b) Random division.

Figure 4 Maps of the scenicness prediction results with 'user dependent random division' and 'fully random division'.

urban, developed scenes through more rural landscapes to natural landscapes and a long tail of nouns associated with the Highlands of Scotland. This long tail also reveals one limitation of our approach, since our natural language processing methods cannot deal with Gaelic, and some misclassified words remained in the list of nouns (e.g. *ruadh* refers to the colour red in Gaelic and is commonly used in toponyms i.e. *Sgurr Ruadh* refers to the Red Peak).

Exploration of the word clouds (Fig. 3) reveals that the scenicness of individual terms sometimes contradicts classic ideas in work on landscape preference. For example, water is commonly associated with scenic landscapes [40, 31], yet in our word clouds it has an average scenicness of only 3–5. On closer examination it becomes apparent that water lies in a word cloud containing many rural terms, and the presence of water is common in such scenes. However, at least in our data, rural as opposed to perceived natural scenes are less highly rated. Thus, treating individual nouns (or terms in general) as predictors of scenicness is difficult, and our word clouds reveal more information about the complex interplay between language and landscape. They further indicate the importance of using language, as opposed to purely data-driven approaches to exploring landscape. Approaches extracting landscape properties using intrinsic landscape qualities from standard spatial datasets and associating these with landscape preferences (e.g. [9, 38]) based on ideas of evolutionary-driven landscape perception [18] are unlikely to capture variation of the nature we observe here. Furthermore, our word clouds are potentially powerful tools for generating datasets containing imagery for use in landscape preference experiments and modelling, since they provide an empirical basis for terms used in selecting candidate images, as opposed to approaches based on introspective reasoning or intrinsic, evolutionary determined preferences (cf. [37]) to generate candidate keywords for querying.

Our model of scenicness, irrespective of training data is able to explain some 52% of the variation in scenicness. This is comparable with typical results in more traditional approaches based on interviews or participatory methods [25], approaches using land cover data [35] and work at a continental scale using social media [37]. Although the explained variance is not strongly influenced by our choice of training data, the total number of grid cells for which average scenicness value can be predicted varies by some 20% from around 7000 cells where individual users are only allowed to be present in either test or training data, to 9000 cells where image descriptions are randomly assigned to test or training data. Furthermore, this variation is strongly spatially autocorrelated, with, for example, a single user having taken some 11000 images in the Lake District National Park, of which ca. 850 were rated in the ScenicOrNot project. Such biases are a typical issue in VGI [13], whose handling requires care. Our results were also sensitive to resolution - finer granularities of model reduced model performance (e.g. at 2.5km we could explain 41% of the variation) and coarser granularities increased model performance (e.g. at 10km we could explain 67% of the variation). These results are not unexpected, since firstly the available training data is reduced as resolution becomes finer and, secondly, a coarser model smooths variation and is thus easier to predict, but conveys less fine grained information at the landscape scale.

Our best model used relatively simple features (800 most frequent unigrams, tf-idf and a dictionary of adjectives associated with landscape). Using bigrams, which might be expected to better capture noun phrases associated with scenic locations (e.g. ‘pleasant landscape’) did not in practice improve model performance, an observation which has been made in other contexts [26]. Verbs of perception appear equally likely to be used in scenic or non-scenic contexts, and were also not useful features in our model.

To our knowledge, our approach is the first attempt to use language to spatially model landscape preference, and it has obvious potential to be combined with other approaches to modelling scenicness based either on user frequentation, physical properties of landscape, or combinations thereof [36, 32, 37].

5 Conclusions and outlook

Our work took advantage of two datasets created by volunteers with very different characteristics. Key to their use in our research were firstly the size and spatial extent of both datasets, and secondly the richness of the textual descriptions associated with Geograph images. Our results demonstrate ways in which VGI and crowdsourcing can allow us to explore questions about how space, and in our case scenicness, is captured through use of language, and demonstrate the potential of such approaches. In particular, we observed:

- clear patterns in the nouns associated with scenicness, suggesting a continuum from heavily developed scenes through more rural to perceived natural scenes. Interpreting and using terms to explain scenicness in isolation is challenging, and we suggest that terms should be analysed in isolation with caution;
- a language-based model can predict some 52% of variance in scenicness, comparable with traditional approaches and state of the art statistical models based on parameters known to correlate with scenicness (e.g. terrain roughness or presence of water). Our approach allows us to capture potentially culturally varying landscape preference through the proxy of language; and
- explained variance was not strongly influenced by the way single users describe landscapes. This makes it unnecessary to restrict the appearance of descriptions of single user either in training or in test datasets.

It is important to note that the approaches we take to modelling scenicness, in contrast to our interpretation of word clouds, essentially use a *bag of words* model, where dependencies between terms are not explicitly modelled. In future research we will explore whether, and how, modelling such dependencies might contribute to our understanding of landscape aesthetics. Importantly, we do not claim that our results are universal, but rather reflect the relationship between landscape and language in a particular cultural setting.

We see this work as an example of the use of textual descriptions to explore culturally determined properties of landscape through language. We also intend to explore the transferability of our results to other user generated content (e.g. Flickr or OpenStreetMap), to other spatial regions and languages (e.g. on mainland Europe) and the impact of including additional spatial data on model performance (e.g. terrain models or land cover data). Furthermore, we see great value in attempting to use the literature to build a taxonomy of scene types, and explore their influence on our model. Such an approach could also take advantage of the "unwritten" parts of our descriptions, for example in terms of the arrangements or presence of objects in a particular image or the relationships between colours through content-based analysis of image content associated with descriptions.

Acknowledgements. We would like to thank Michele Volpi for the fruitful discussions and his and the reviewers helpful comments. We would also like to gratefully acknowledge contributors to ScenicOrNot (Open Database Licence) and Geograph British Isles (Creative Commons Attribution-ShareAlike 2.5 Licence).

References

1 J. Appleton. Prospect and Refuges Revisited. In J. L. Nasar, editor, *Environmental Aesthetics: Theory, Research and Applications*, pages 27–44. Cambridge University Press, 1988.

2 J. A. Benfield, P. A. Bell, L. J. Troup, and N. C. Soderstrom. Aesthetic and affective effects of vocal and traffic noise on natural landscape assessment. *Journal of Environmental Psychology*, 30(1):103–111, 2010. doi:10.1016/j.jenvp.2009.10.002.

3 S. Casalegno, R. Inger, C. DeSilvey, and K. J. Gaston. Spatial Covariance between Aesthetic Value & Other Ecosystem Services. *PLOS ONE*, 8(6), 2013. doi:10.1371/journal.pone.0068437.

4 A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision*, 7(2-3):81–227, 2012.

5 A. Dunkel. Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and Urban Planning*, 142:173–186, 2015. doi:10.1016/j.landurbplan.2015.02.022.

6 A. J. Edwardes and R. S. Purves. A theoretical grounding for semantic descriptions of place. *Proceedings of the 7th international conference on Web and wireless geographical information systems*, pages 106–120, 2007. doi:10.1007/978-3-540-76925-5_8.

7 S. Elwood, M. F. Goodchild, and D. Z. Sui. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, 102(3):571–590, 2012. doi:10.1080/00045608.2011.595657.

8 P. Fisher and D. J. Unwin. Re-presenting Geographical Information Systems. In Fisher, Peter and Unwin, David J., editor, *Re-presenting GIS*, pages 1–14. Wiley Sons London, London, 2005.

9 S. Frank, Ch. Fürst, A. Witt, L. Koschke, and F. Makeschin. Making use of the ecosystem services concept in regional planning-trade-offs from reducing water erosion. *Landscape Ecology*, pages 1–15, 2014. doi:10.1007/s10980-014-9992-3.

10 F. Girardin, J. Blat, F. Calabrese, F. Dal Fiore, and C. Ratti. Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing*, 7(4):36–44, 2008. doi:10.1109/MPRV.2008.71.

11 G. Gliozzo, N. Pettorelli, and M. Haklay. Using crowdsourced imagery to detect cultural ecosystem services: a case study in South Wales, UK. *Ecology and Society*, 21(3), 2016. doi:10.5751/ES-08436-210306.

12 M. F. Goodchild. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007. doi:10.1007/s10708-007-9111-y.

13 M. Haklay. Why is participation inequality important? In C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, and R. Purves, editors, *European Handbook of Crowdsourced Geographic Information*, pages 35–44. Ubiquity Press, London, 2016.

14 V. Hatzivassiloglou and J. M. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. *Proceedings of the 18th conference on Computational linguistics-Volume 1*, pages 299–305, 2000. doi:10.3115/990820.990864.

15 M. Hunziker, M. Buchecker, and T. Hartig. Space and Place – Two Aspects of the Human-landscape Relationship. *Challenges for Landscape Research*, pages 47–62, 2007. doi:10.1007/978-1-4020-4436-6_5.

16 A. Jenkins, A. Croitoru, A. T. Crooks, and A. Stefanidis. Crowdsourcing a Collective Sense of Place. *Plos One*, 11(4):e0152932, 2016. doi:10.1371/journal.pone.0152932.

17 X. Junge, B. Schüpbach, T. Walter, B. Schmid, and P. Lindemann-Matthies. Aesthetic quality of agricultural landscape elements in different seasonal stages in Switzerland. *Landscape and Urban Planning*, 133:67–77, 2015. doi:10.1016/j.landurbplan.2014.09.010.

18 R. Kaplan and S. Kaplan. *The experience of nature: a psychological perspective*. Cambridge University Press, 1989.

19 A. Lothian. Landscape and the philosophy of aesthetics: Is landscape quality inherent in the landscape or in the eye of the beholder? *Landscape and Urban Planning*, 44(4):177–198, 1999. doi:10.1016/S0169-2046(99)00019-5.

20 A. Lothian. Scenic perceptions of the visual effects of wind farms on South Australian landscapes. *Geographical Research*, 46(2):196–207, 2008. doi:10.1111/j.1745-5871.2008.00510.x.

21 Ch. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts, 1999. doi:10.1145/601858.601867.

22 D. M. Mark, A. G. Turk, N. Burenhult, and D. Stea, editors. *Landscape in language. Transdisciplinary perspectives*. John Benjamins, Amsterdam/Philadelphia, 2011.

23 J. Mascaro, G. P. Asner, D. E. Knapp, T. Kennedy-Bowdoin, R. E. Martin, Ch. Anderson, M. Higgins, and K. D. Chadwick. A tale of two "Forests": Random Forest machine learning aids tropical Forest carbon mapping. *PLoS ONE*, 9(1):12–16, 2014. doi:10.1371/journal.pone.0085993.

24 J. L. Nasar. *Environmental Aesthetics: Theory, research and Application*. Cambridge edition, 1992.

25 J. F. Palmer. Using spatial metrics to predict scenic perception in a changing landscape: Dennis, Massachusetts. *Landscape and Urban Planning*, 69(2-3):201–218, 2004. doi:10.1016/j.landurbplan.2003.08.010.

26 B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. *Empirical Methods in Natural Language Processing (EMNLP)*, 10(July):79–86, 2002. doi:10.3115/1118693.1118704.

27 T. Plieninger, S. Dijks, E. Oteros-Rozas, and C. Bieling. Assessing, mapping, and quantifying cultural ecosystem services at community level. *Land Use Policy*, 33:118–129, 2013. doi:10.1016/j.landusepol.2012.12.013.

28 R. S. Purves, A. J. Edwardes, and J. Wood. Describing place through user generated content. *First Monday. Peer-reviewed journal on the internet*, 16(9), 2011. doi:10.5210/fm.v16i9.3710.

29 T. Rattenbury and M. Naaman. Methods for extracting place semantics from Flickr tags. *ACM Transactions on the Web*, 3(1):1–30, 2009. doi:10.1145/1462148.1462149.

30 D. R. Richards and D. A. Friess. A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators*, 53:187–195, 2015. doi:10.1016/j.ecolind.2015.01.034.

31 U. Schirpke, E. Tasser, and U. Tappeiner. Predicting scenic beauty of mountain regions. *Landscape and Urban Planning*, 111(1):1–12, 2013. doi:10.1016/j.landurbplan.2012.11.010.

32 C. I. Seresinhe, H. S. Moat, and T. Preis. Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science*, page 026581351668730, 2017. doi:10.1177/0265813516687302.

33 C. I. Seresinhe, T. Preis, and H. S. Moat. Quantifying the Impact of Scenic Environments on Health. *Scientific Reports*, 5(Article number 16899):1–9, 2015. doi:10.1038/srep16899.

34 B. Smith and D. M. Mark. Geographical categories: an ontological investigation. *International Journal of Geographical Information Science*, 15(7):591–612, 2001. doi:10.1080/13658810110061199.

35 B. Stadler, R. Purves, and M. Tomko. Exploring the Relationship Between Land Cover and Subjective Evaluation of Scenic Beauty through User Generated Content. In *Proceedings of the 25th International Cartographic Conference*, Paris, 2011. doi:10.5167/uzh-52945.

36 P. Tenerelli, U. Demšar, and S. Luque. Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators*, 64:237–248, 2016. doi:10.1016/j.ecolind.2015.12.042.

37 B. T. van Zanten, D. B. van Berkel, R. K. Meetemeyer, J. W. Smith, K. F. Tieskens, and P. H. Verburg. Continental scale quatification of landscape values using social media data. *Proceedings of the National Academy of Sciences*, pages 1–7, 2016. doi:10.1073/pnas.1614158113.

38 B. T. van Zanten, P. H. Verburg, M. J. Koetse, and P. J. H. van Beukering. Preferences for European agrarian landscapes: A meta-analysis of case studies. *Landscape and Urban Planning*, 132:89–101, 2014. doi:10.1016/j.landurbplan.2014.08.012.

39 A. Viberg. The verbs of perception: a typological study. *Linguistics*, 21(1):123–162, 2009. doi:10.1515/ling.1983.21.1.123.

40 B. Yang and T. J. Brown. A Cross-Cultural Comparison of Preferences for Landscape Styles and Landscape Elements. *Environment and Behavior*, 24(4):471–507, 1992. doi:10.1177/0013916592244003.